Spectral Compression for Remote Sensing Using Principal Component Analysis

Vijay Natraj

Jet Propulsion Laboratory, California Institute of Technology

Yuk Lunch Seminar
June 26, 2019

Acknowledgements

- Xun Jiang, University of Houston
- Xianglei Huang, University of Michigan
- Run-Lie Shia, Caltech
- Yuk Yung, Caltech
- Robert Spurr, RT Solutions, Inc.
- Pushkar Kopparla, University of Tokyo
- David Crisp, Jet Propulsion Laboratory, Caltech
- Peter Somkuti, Colorado State University
- Chao Liu, Nanjing University of Information Science and Technology
- Xu Liu, NASA Langley Research Center
- … and many others

Rationale / Need for Speed

- **RT models required for generating simulated radiances from satellite, ground-based and other platforms**

- **In retrieval applications, RT models also needed to calculate Jacobians**

- **OSSEs and climate models require massive RT modeling over wide spectral ranges**

- **RT calculations computationally expensive**

- **New generation LEO and GEO satellites expected to generate data at rates current computing power is unlikely to match**

Overview of Existing Techniques I

- **Correlated-k/Exponential Sum Fitting of Transmittances**
  - Widely used for atmospheric heating/cooling rate calculations
  - Assume that optical properties spectrally correlated at all points along optical path
  - Only valid for homogeneous, isobaric, isothermal atmospheres
  - Loss of correlation can introduce significant radiance errors

Overview of Existing Techniques II

- **Spectral Mapping**
  - **No assumption about spectral correlation along optical path**
  - **Perform level-by-level comparison of monochromatic atmospheric and surface optical properties**
  - **Combine only spectral regions that remain similar at all points along inhomogeneous optical path**
  - **Fine binning required to achieve high RT calculation accuracy**
  - **Coarse binning provides significant reduction in radiance accuracy**

Overview of Existing Techniques III

- **Asymptotic Methods**
  - **Limited to study of semi-infinite media (e.g. optically thick clouds)**
- **Low Orders of Scattering**
  - **Restricted to study of optically thin atmospheres**
- **Others**
  - **Usefulness only proven for narrow spectral regions**
  - **Many of these techniques rely on finite differences**

Why is RT Computationally Expensive?

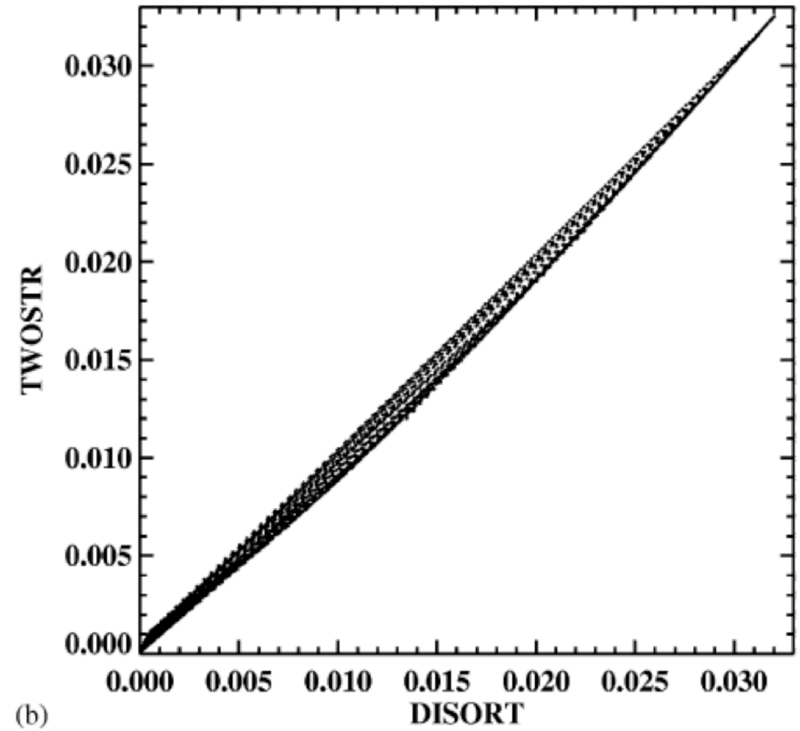
- **Wavelengths**
  - **Spectral points where radiances must be evaluated**

- **Angles**
  - **Computational quadrature angles (“streams”) at each spectral point**

- **A good solution should address BOTH the above considerations.**

Angles

- **Separation of single and multiple scattering**
  - **Large number (sometimes >> 100) of streams required to resolve accurately anisotropy due to scattering**
  - **Computational burden ~O(M^3)**
  - **Large part of anisotropy of radiation field captured by single scattering**
  - **Single scattering calculation computationally efficient**



Natraj et al., 2005

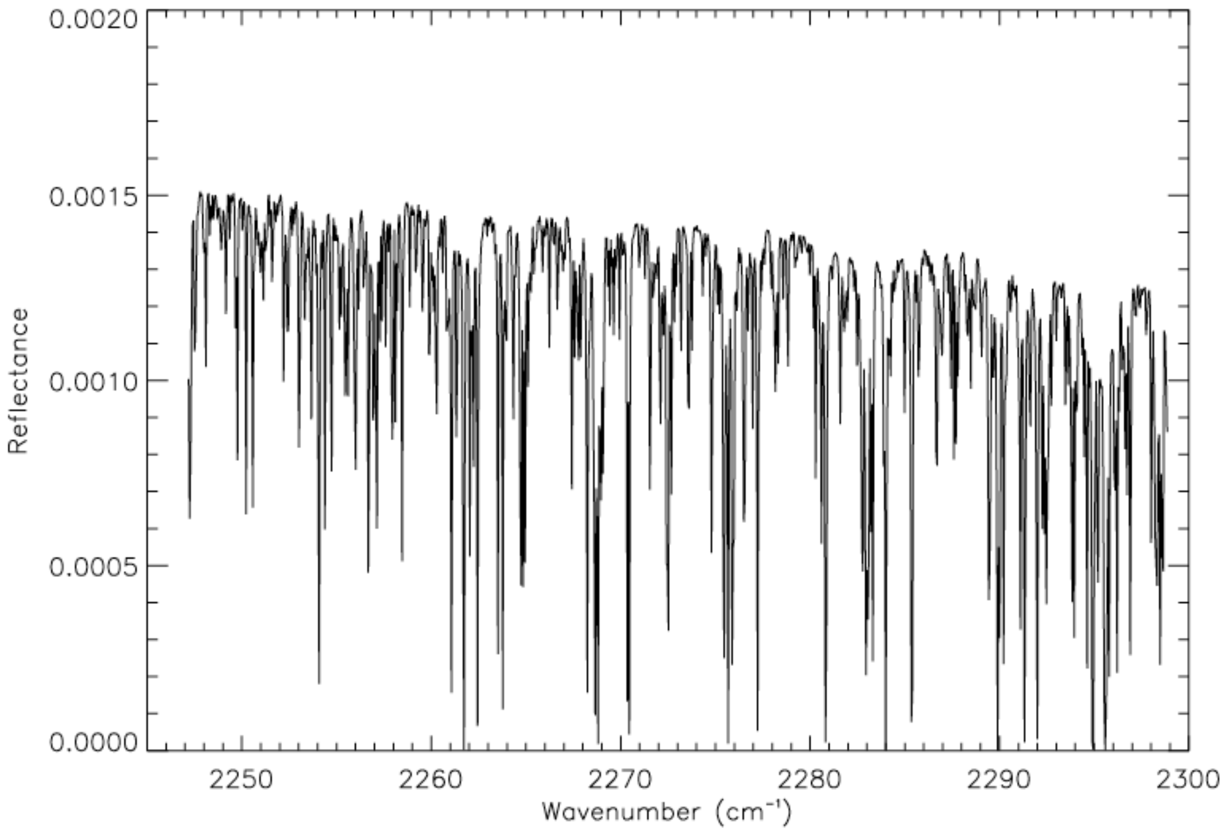
- **Correlation between multi-stream and two-stream calculations**

$$I_{PCA}(\lambda) = [I_{2S}(\lambda)]C(\lambda) + \bar{I}_{FO}(\lambda)$$

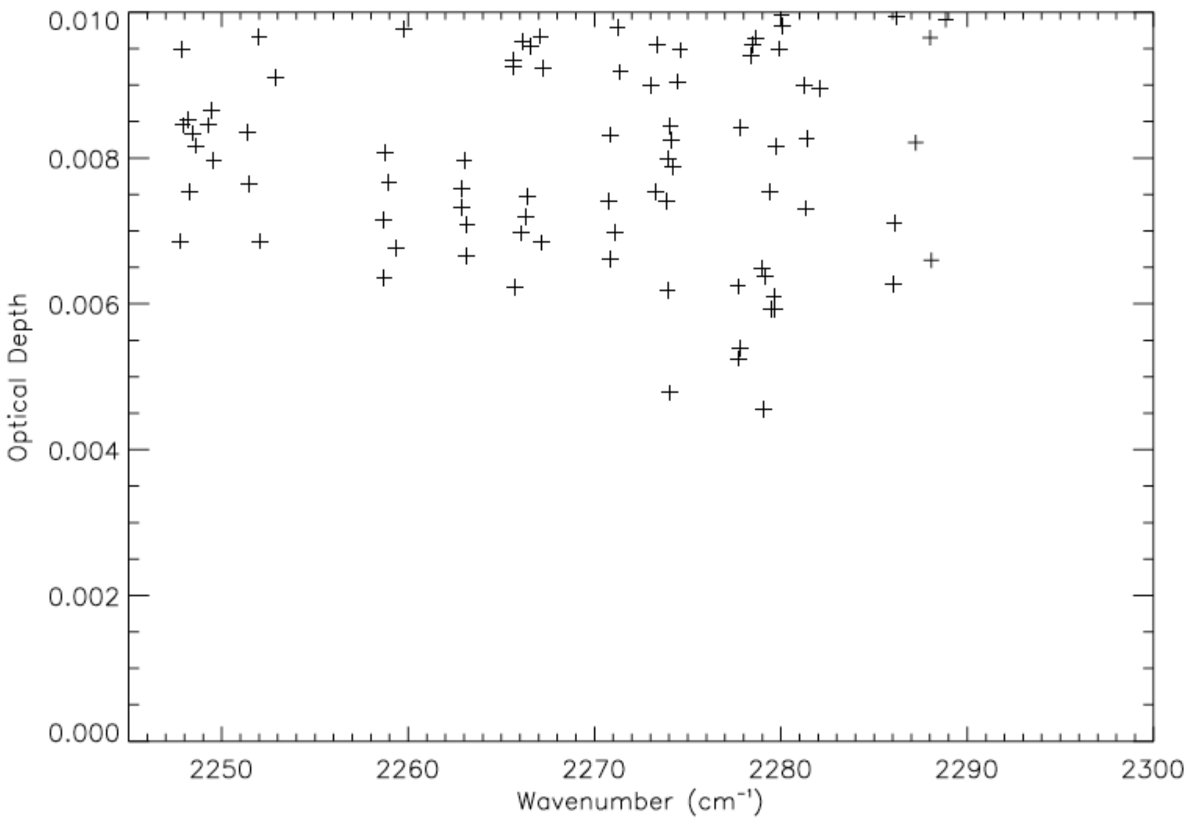
Wavelengths

- **Spectral binning**
- **Eigenvalue problem solution**
- **Radiance back-mapping**

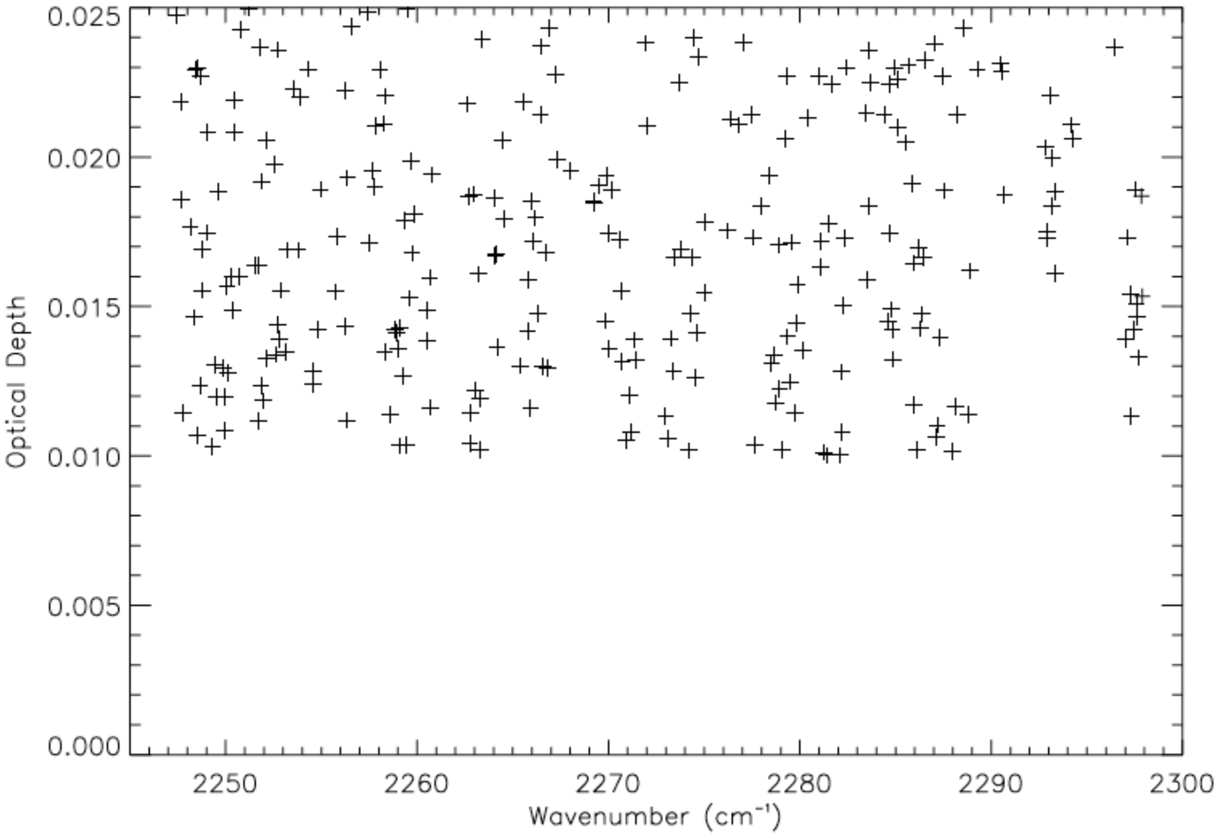
Spectral Binning



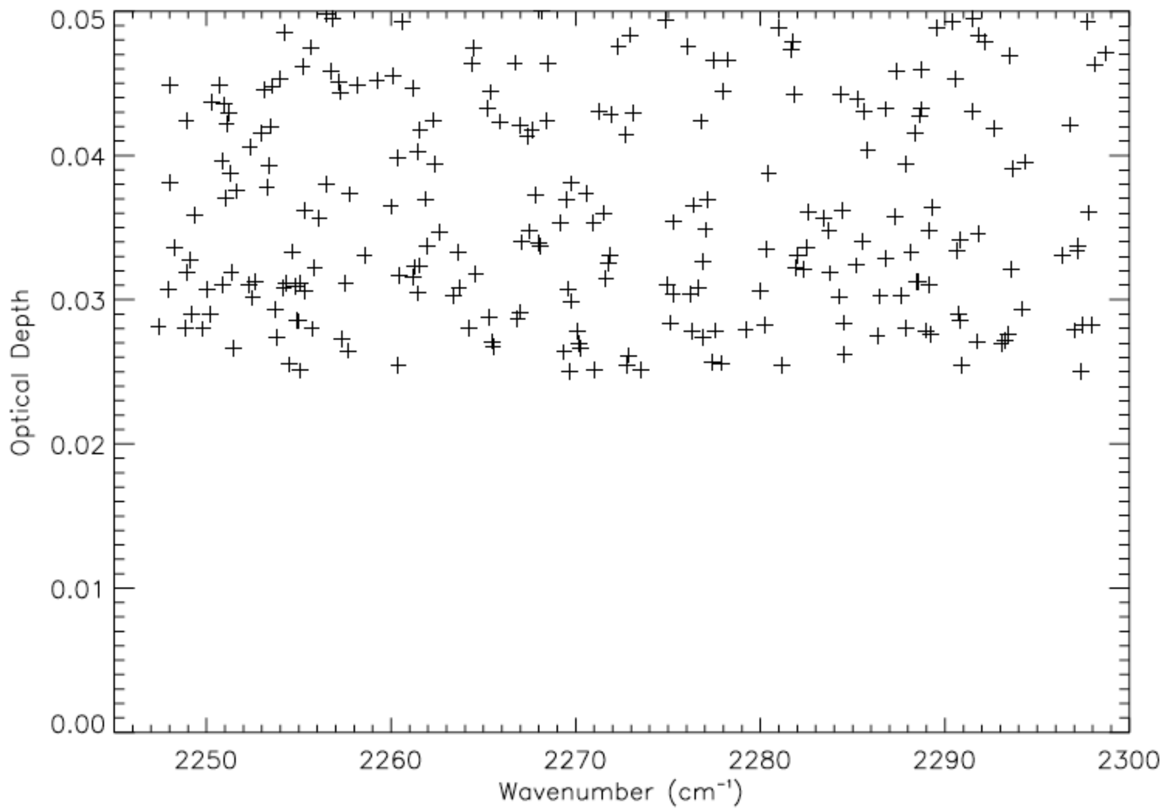
Spectral Binning



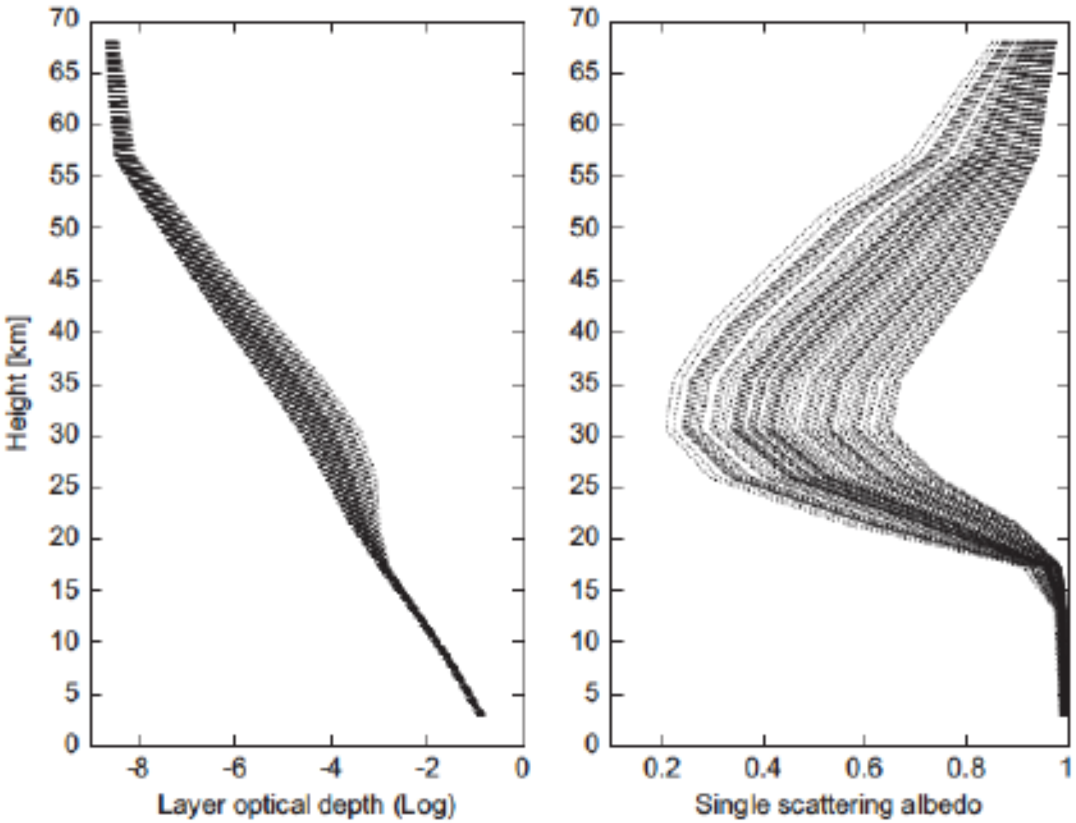
Spectral Binning



Spectral Binning



RT Data Set



Height [km]
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
-8
-6
-4
-2
0
Layer optical depth (Log)
0.2
0.4
0.6
0.8
1
Single scattering albedo

Eigenvalue Problem Solution: PCA

- **Data Set**
  - Optical properties in *M* atmospheric layers at *N* wavelengths
- **Empirical Orthogonal Functions (EOFs)**
  - Eigenvectors of covariance matrix of detrended (mean removed) data set
  - New basis to represent original data
  - No loss of information
- **Principal Component Scores/Weights**
  - Projection of original data set onto EOFs

Eigenvalue Problem Solution: PCA

- **PCA is an orthogonal transformation**
- **EOFs uncorrelated (original data set strongly correlated)**
- **EOFs sorted in order of decreasing variance accounted for**
- **First few (typically <= 4) EOFs capture > 99.99% of variance**
- **PCA gives insight into variability patterns in data sets**

Eigenvalue Problem Solution: PCA

- **Mean and EOFs define much smaller set of PCA-projected optical states compared to original data set**

- **Multiple scattering (MS) simulations performed only on reduced set**

- **Fast two-stream model approximates MS contribution at each spectral point**

- **Correction factors developed based on RT calculations at PCA-projected optical states**

- **Single scatter calculations performed at each spectral point**

Radiance Back-Mapping

- Radiances for mean bin values: I_{exact}, I_{2S}

$$I_d = \ln(I_{exact}/I_{2S})$$

- EOF-perturbed ratios: I_d^+, I_d^-

- First and second order differences

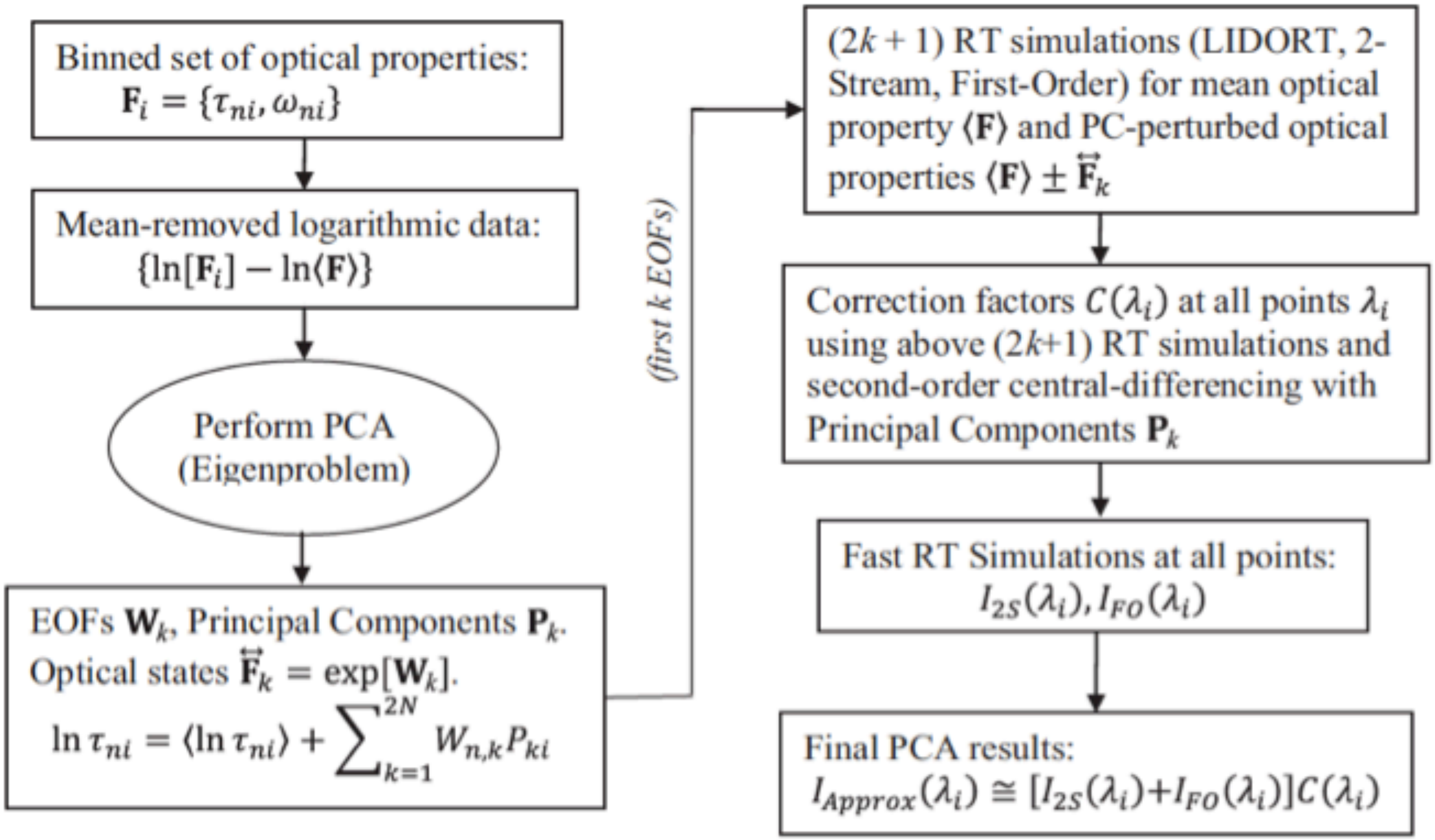
$$\delta I_k = \frac{I_d^+ - I_d^-}{2}$$

$$\delta^2 I_k = I_d^+ + I_d^- - 2I_d$$

- Corrected MS radiance

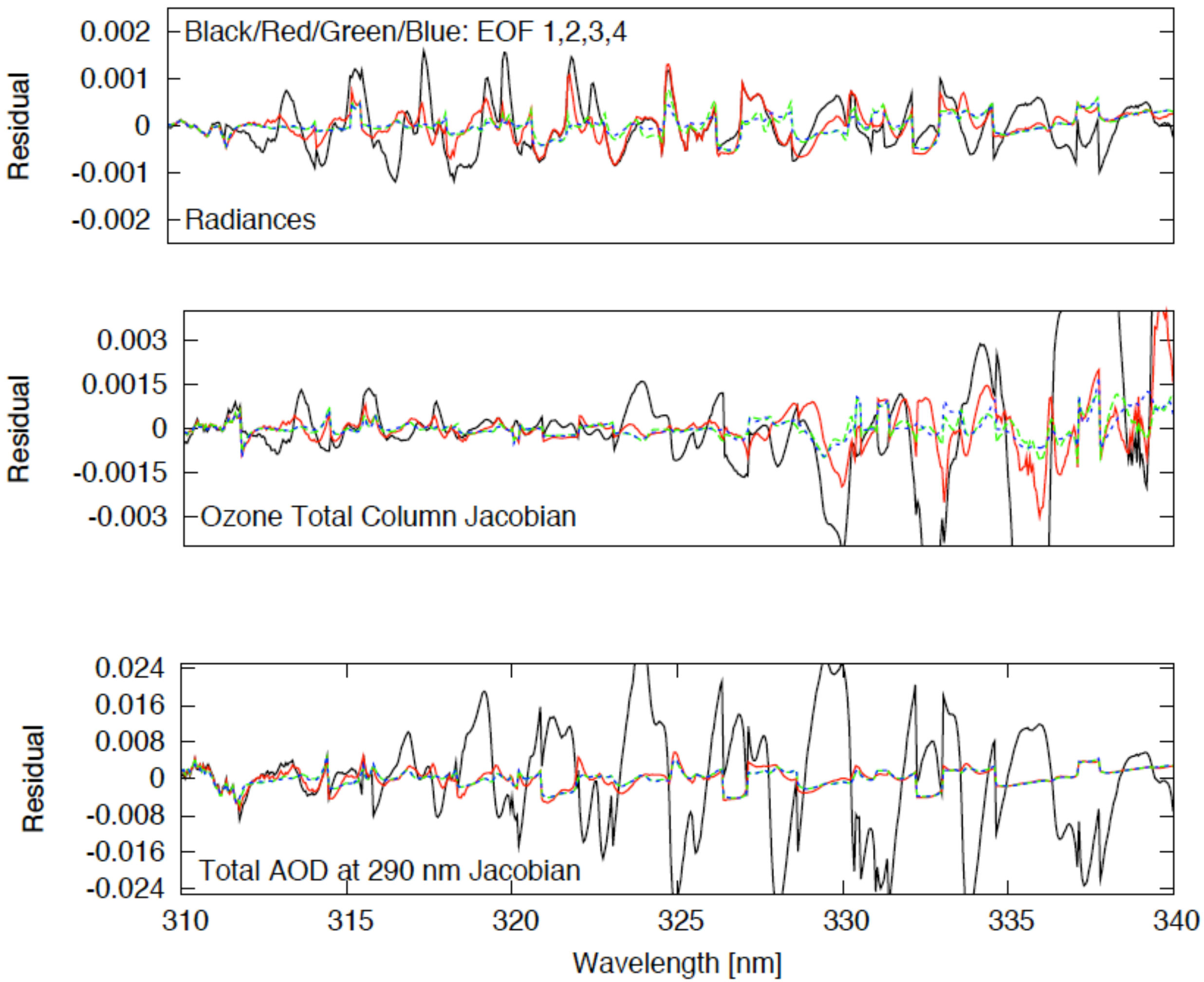
$$I_l = I_l^{2S} \exp\left[I_d + \sum_{k=1}^{4} \delta I_k P_{k,l} + \frac{1}{2}\sum_{k=1}^{4} \delta^2 I_k P_{k,l}^2\right]$$

PCA Flowchart



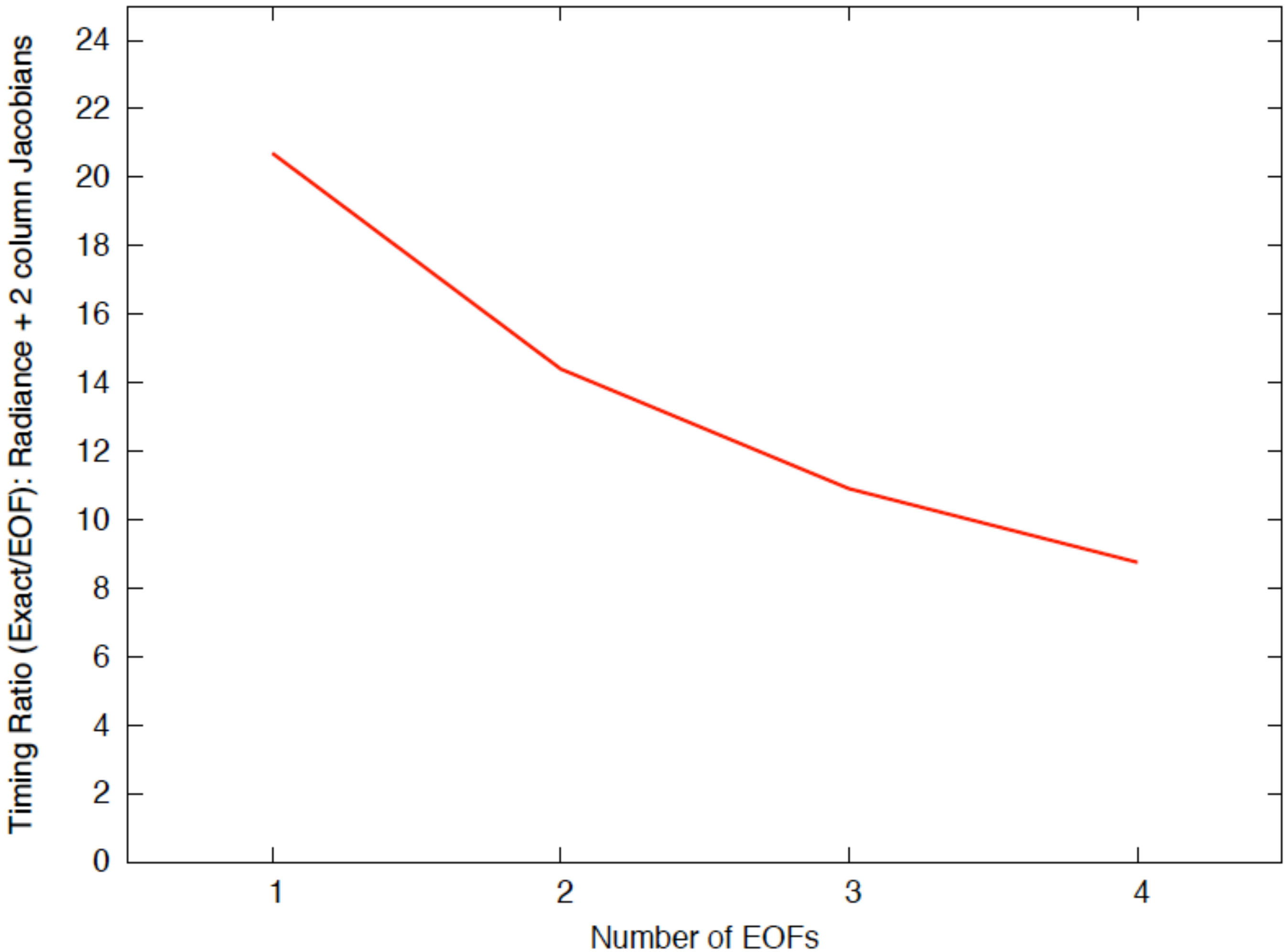
Spurr et al., 2013

UV Ozone Application



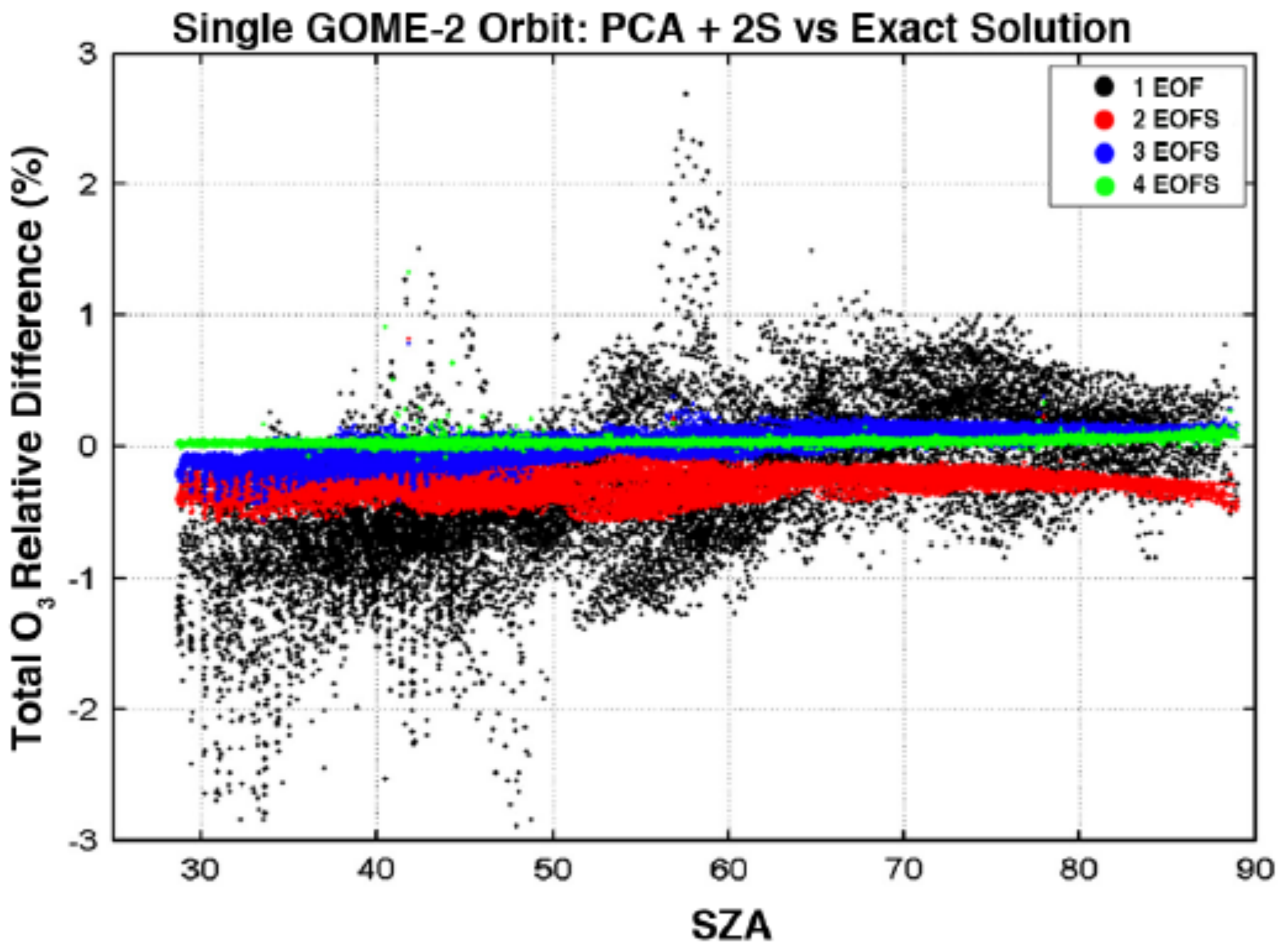
Spurr et al., 2013

Speed Enhancement



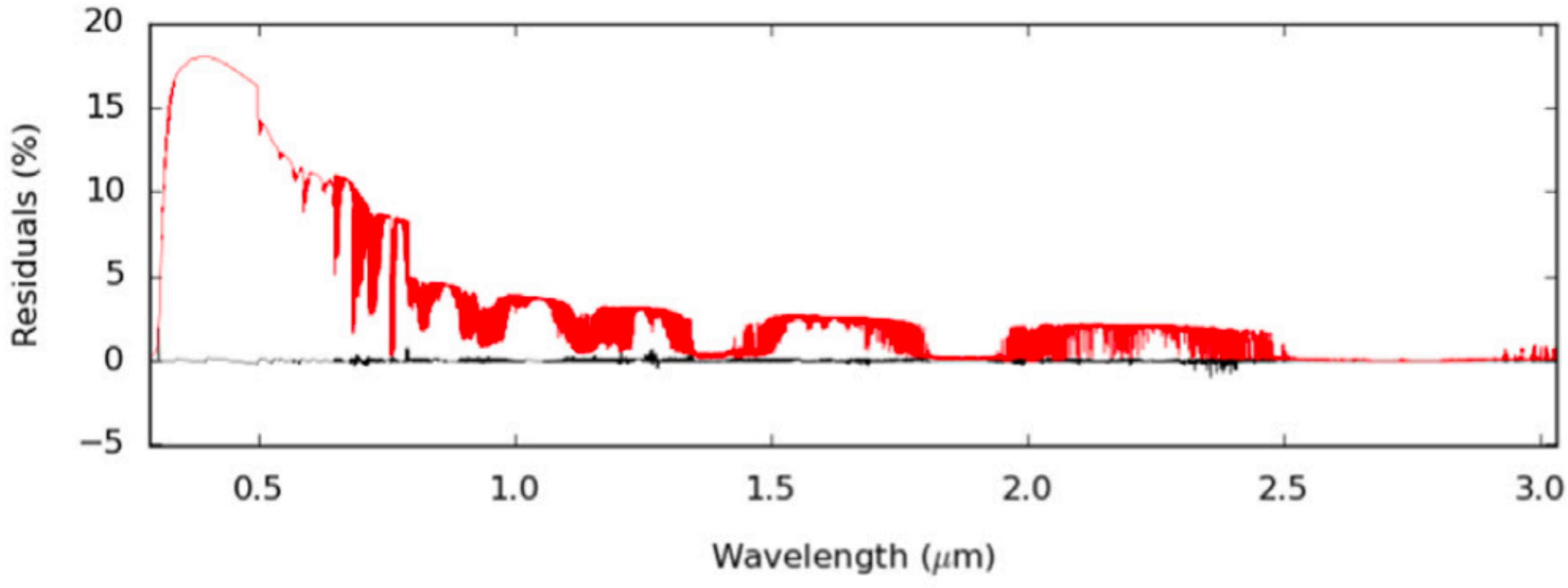
Spurr et al., 2013

GOME-2 Retrievals



Spurr et al., 2013

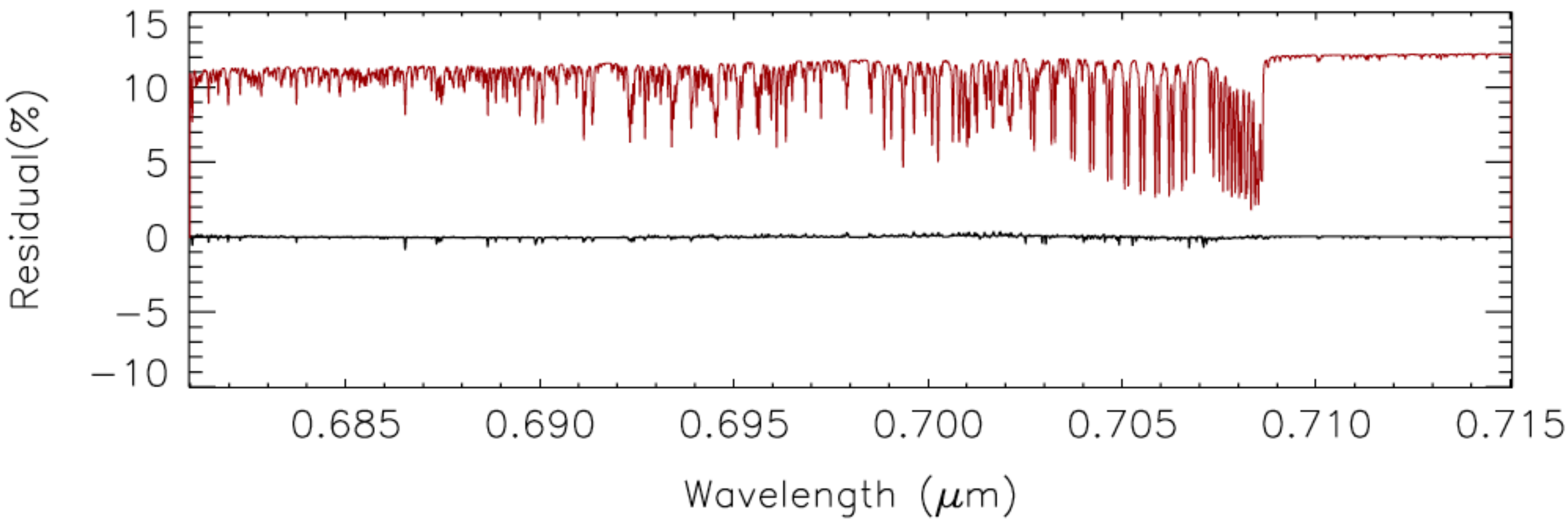
Broadband Radiances



Red: Two-stream; Black: PCA

Kopparla et al., 2016

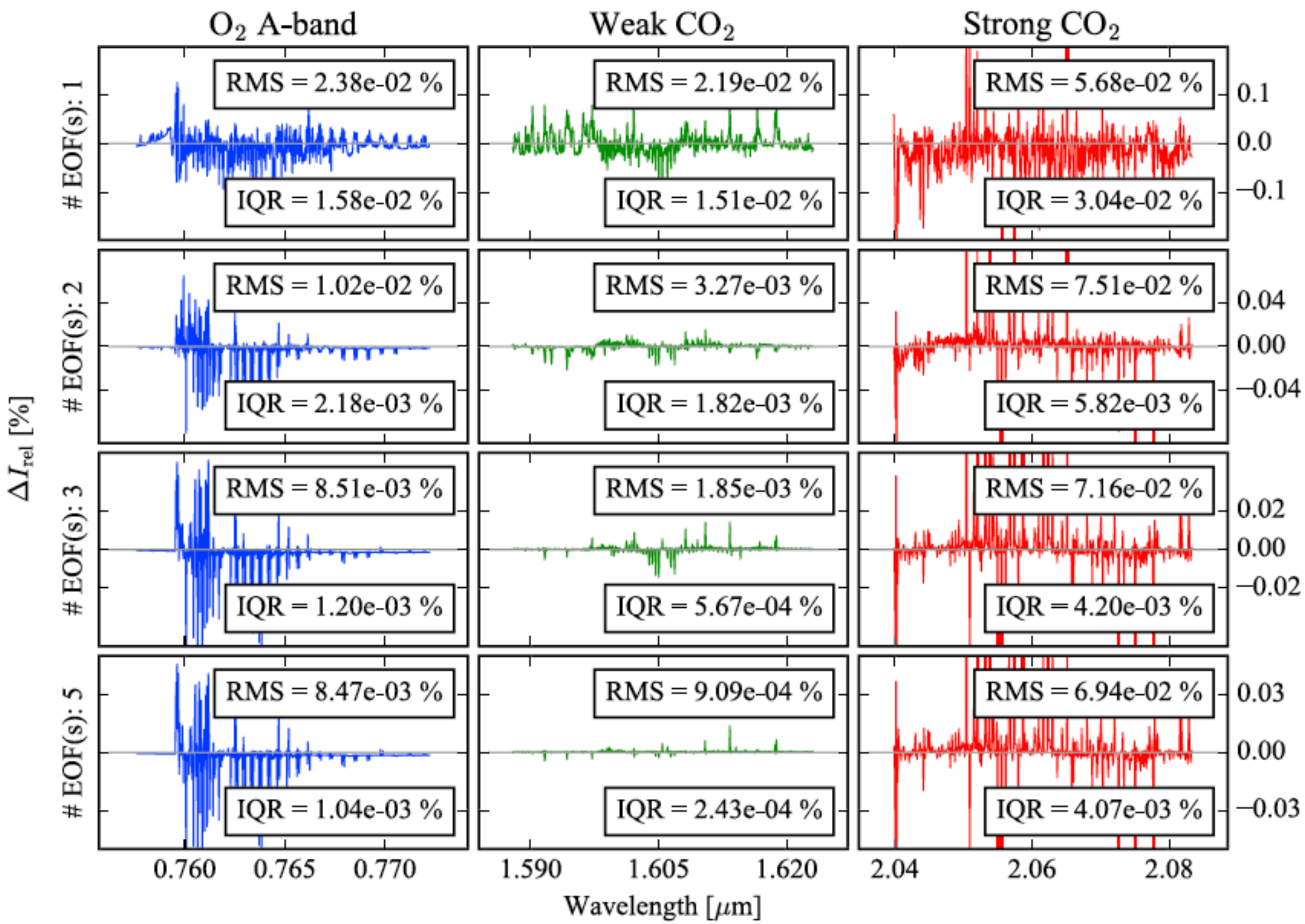
Zoomed in



Brown: Two-stream; Black: PCA

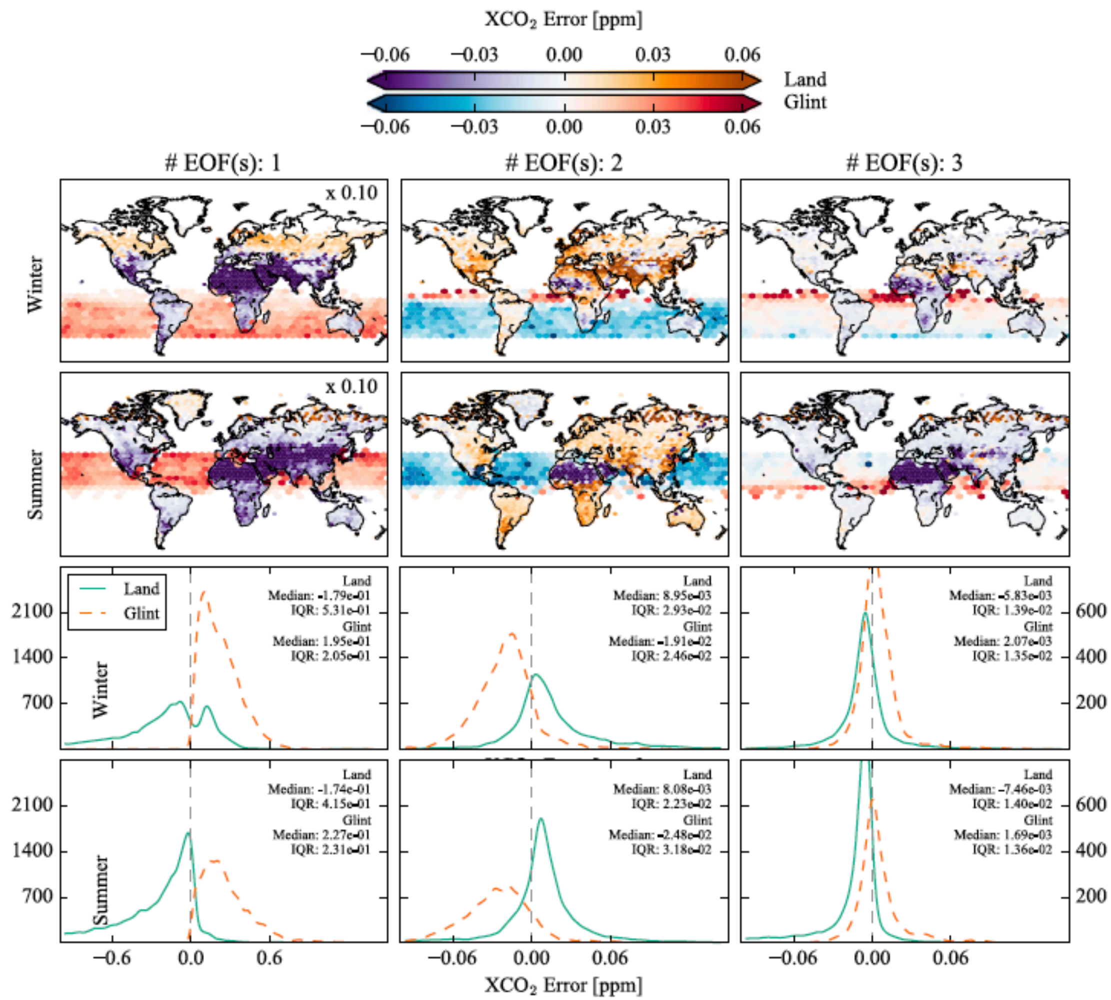
Kopparla et al., 2016

OCO-2 Retrievals



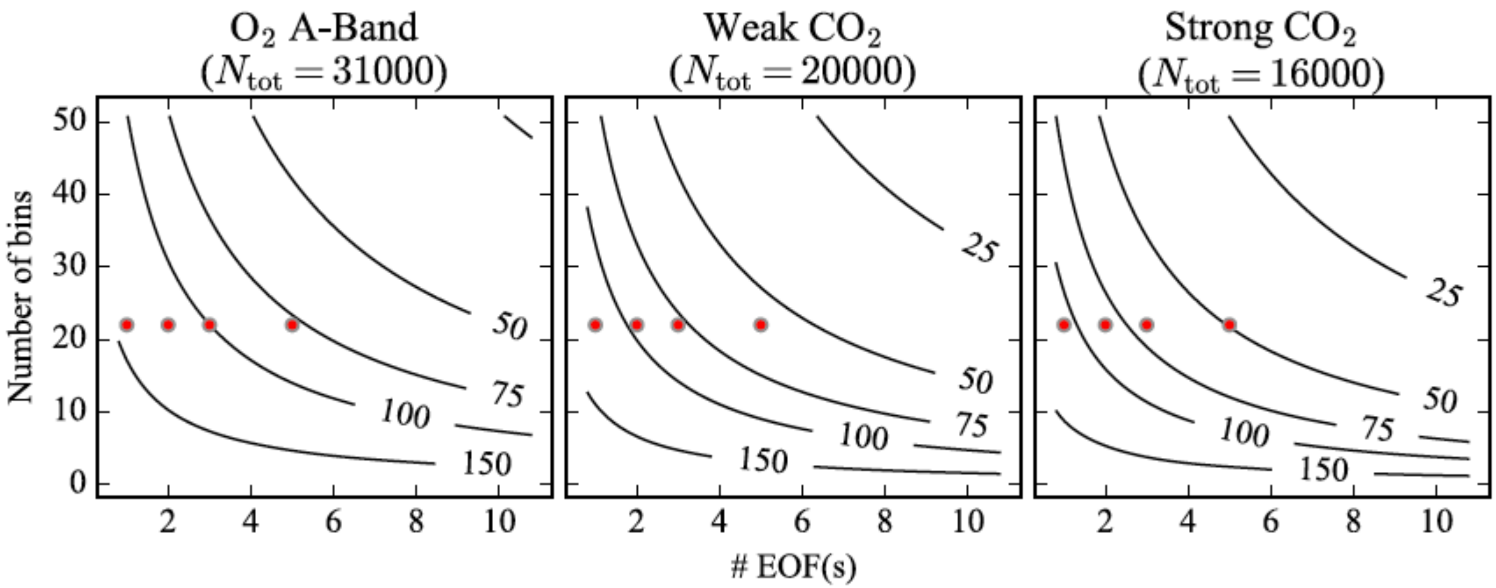
Somkuti et al., 2017

CO_2 Retrieval Errors



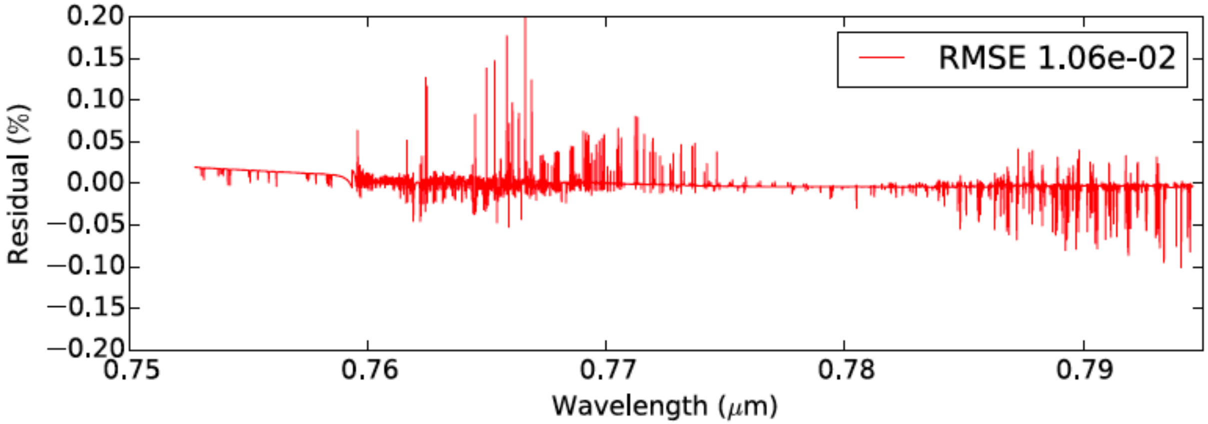
Somkuti et al., 2017

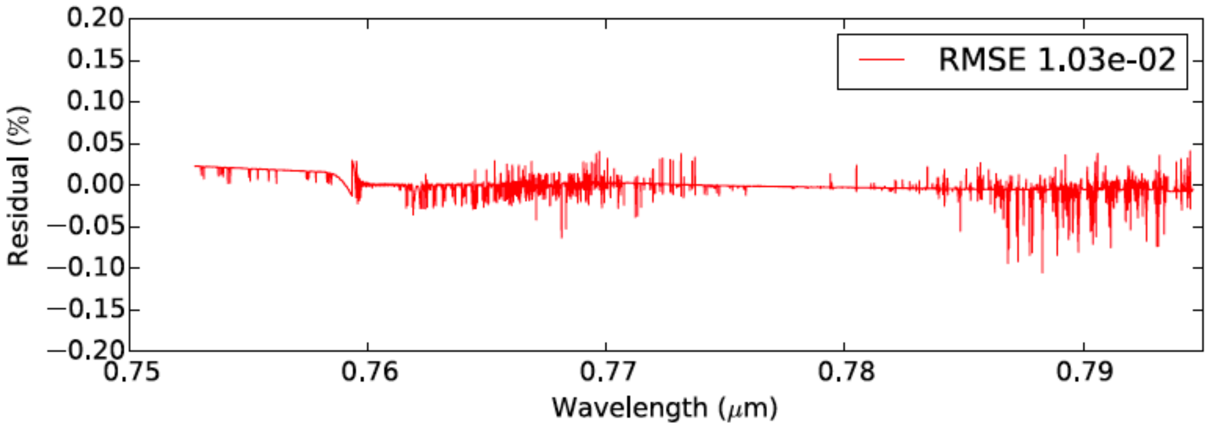
Speed-Up



Somkuti et al., 2017

Vertical Grid

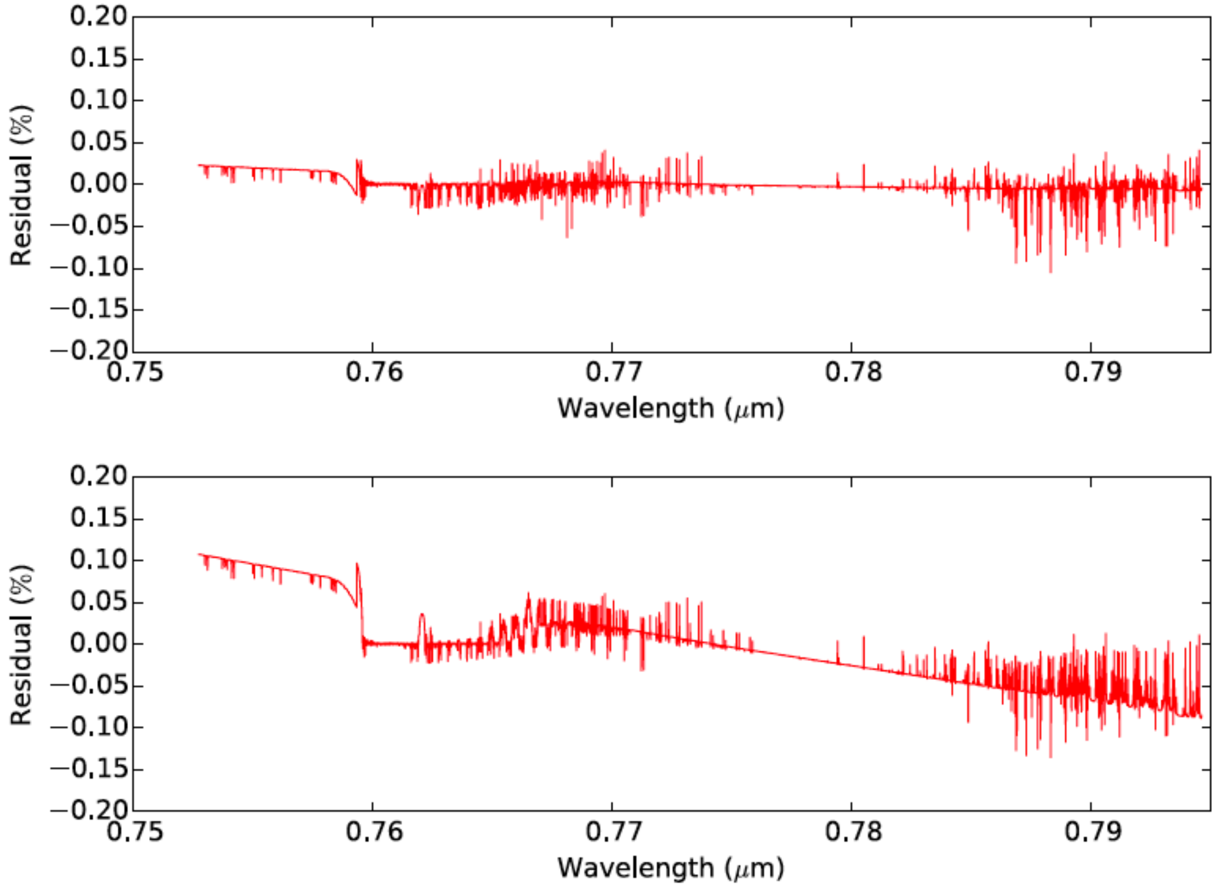




Top: Arbitrary grid; Bottom: Equal pressure thickness grid

Kopparla et al., 2017

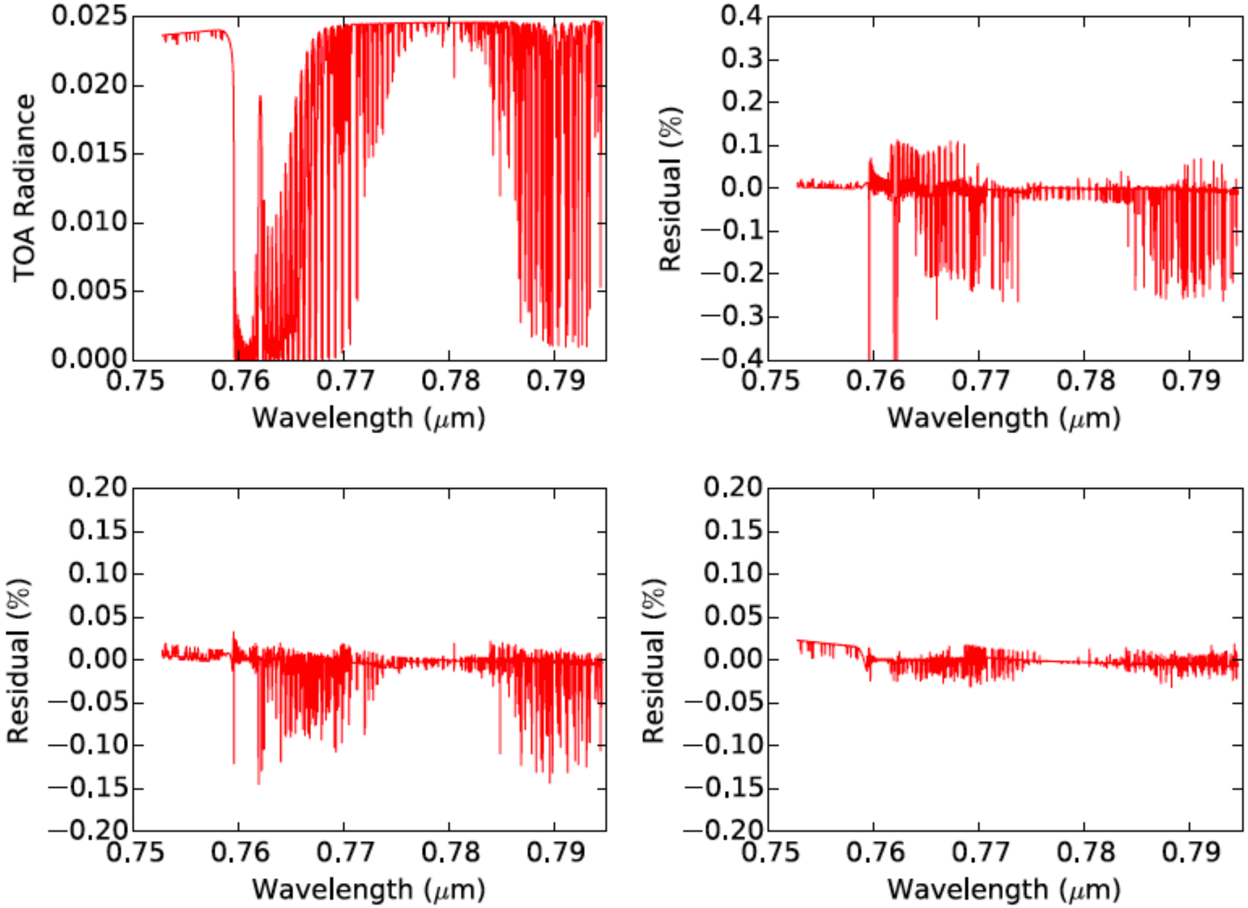
Aerosols in the PCA



Bottom: Bin averaging for aerosol properties; Top: Aerosols in the PCA

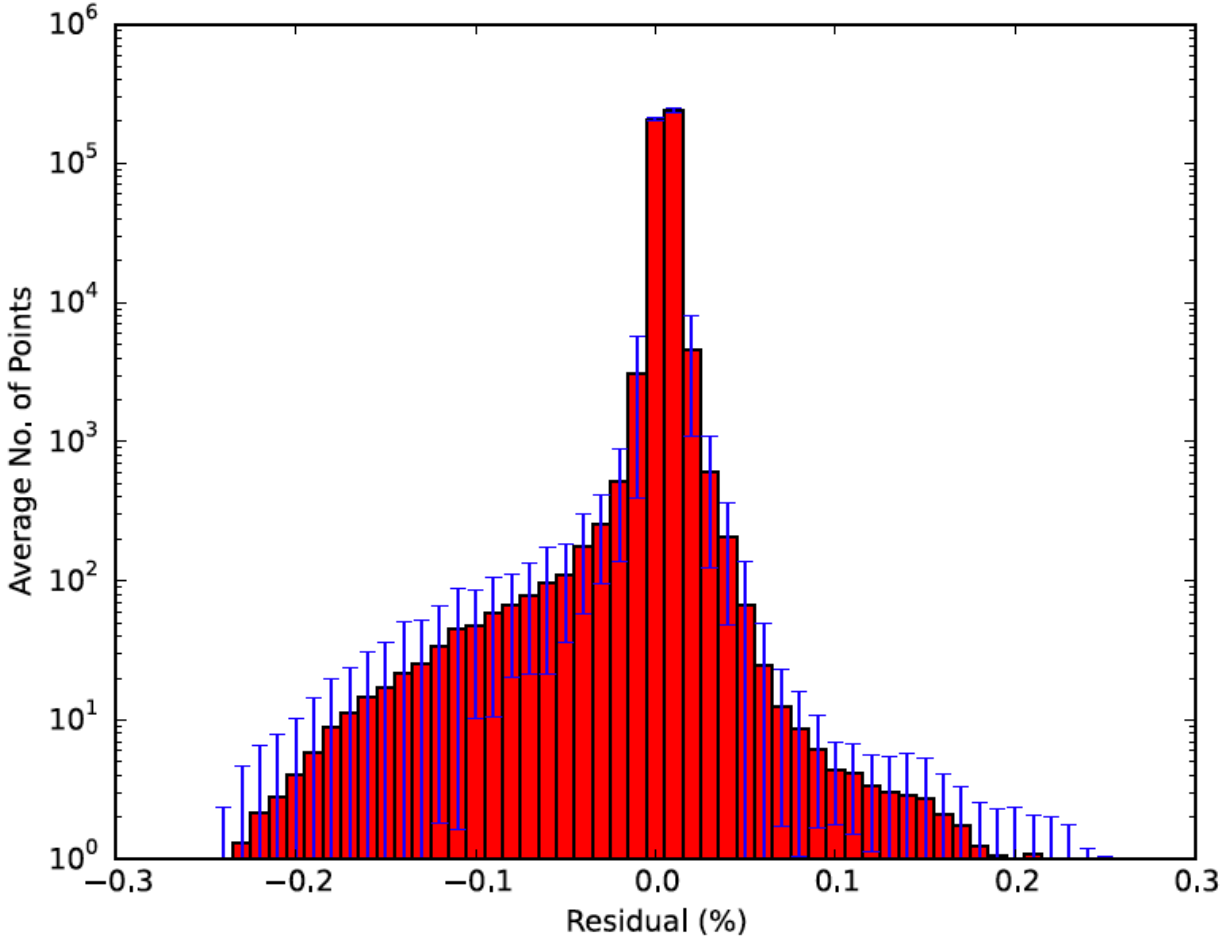
Kopparla et al., 2017

Binning Schemes



Kopparla et al., 2017

Distribution of Residuals



Kopparla et al., 2017

Radiance-Based PCA (Liu, X., et al.)

- **# independent pieces of information << # channels**
- **PCA performed on channel radiances**
- **Large number of atmospheric profiles used to generate a matrix of spectral channel radiances (done offline)**
- **PCA produces EOFs (which are scenario independent)**
- **PC scores for specific scenario obtained using monochromatic calculations at selected wavelengths**

Unified PCA

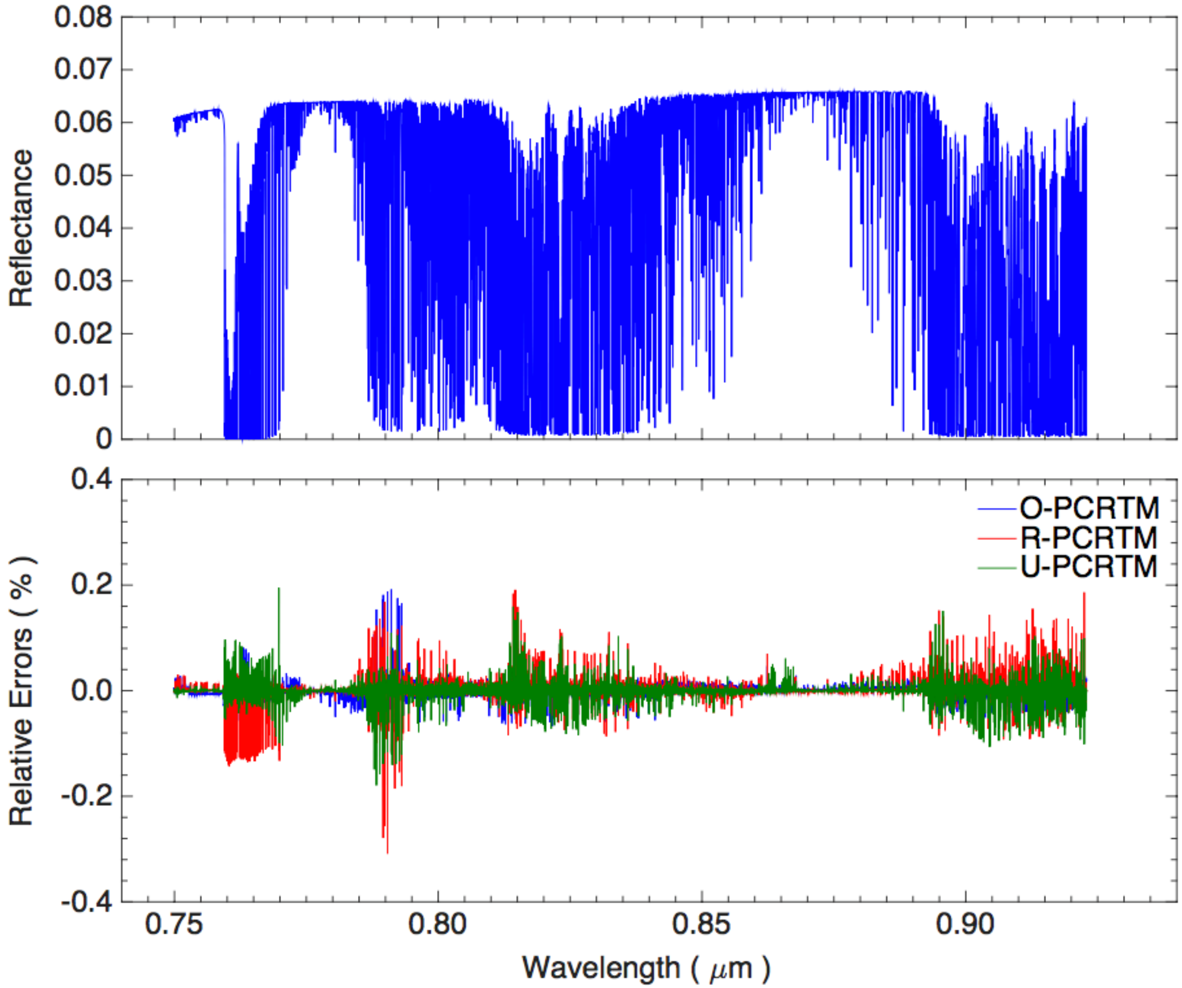
- **Optical-PCA minimizes number of RT calculations per instrument channel**

- **Radiance-PCA minimizes number of instrument channels for which RT calculations are performed**

- **Two techniques are complementary**

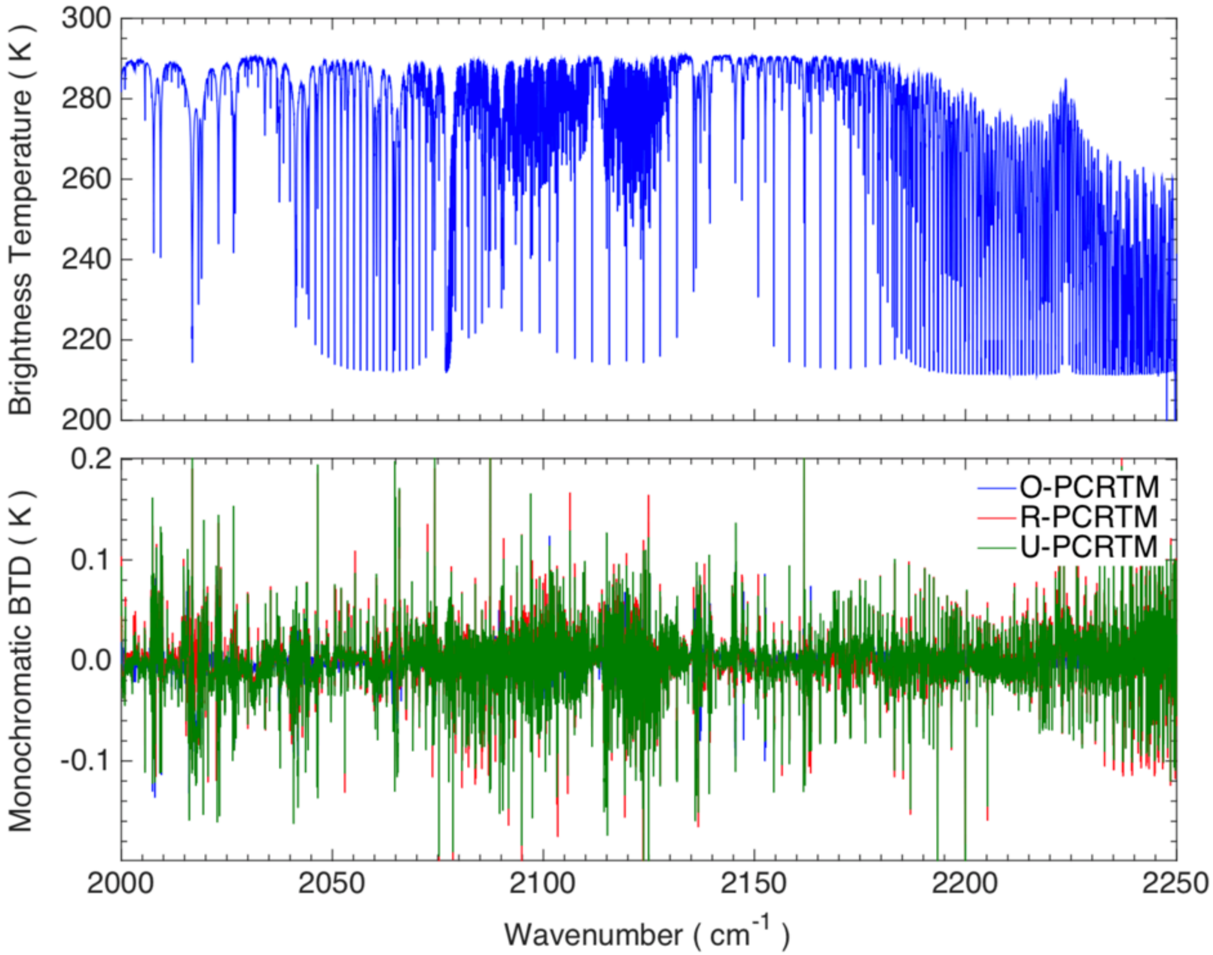
- **Unified PCA model would combine advantages of both methods**

Unified PCA Example 1



Liu et al., 2019

Unified PCA Example 2



Liu et al., 2019

Future Work

- **Vertical layering**
- **Spectral sampling**
- **Binning**
- **Polarization**
- **Unified PCA**
- **Remote sensing retrievals / climate models**